GRM: Large Gaussian Reconstruction Model for Efficient 3D Reconstruction and Generation

Abstract

We introduce GRM, a large-scale reconstructor capable of recovering a 3D asset from sparse-view images in around 0.1s. GRM is a feed-forward transformer-based model that efficiently incorporates multi-view information to translate the input pixels into pixel-aligned Gaussians, which are unprojected to create a set of densely distributed 3D Gaussians representing a scene. Together, our transformer architecture and the use of 3D Gaussians unlock a scalable and efficient reconstruction framework. Extensive experimental results demonstrate the superiority of our method over alternatives regarding both reconstruction quality and efficiency. We also showcase the potential of GRM in generative tasks, i.e., text-to-3D and image-to-3D, by integrating it with existing multi-view diffusion models. Our project website is at: https://justimyhxu.github.io/projects/grm/.